## ORIGINAL ARTICLE

# Rapid genomic DNA changes in allotetraploid fish hybrids

J Wang[1,3], LH Ye[1,3], QZ Liu[2,3], LY Peng[1], W Liu[1], XG Yi[1], YD Wang[1], J Xiao[1], K Xu[1], FZ Hu[1], L Ren[1], M Tao[1], C Zhang[1], Y Liu[1], YH Hong[2] and SJ Liu[1]

Rapid genomic change has been demonstrated in several allopolyploid plant systems; however, few studies focused on animals. We addressed this issue using an allotetraploid lineage (4nAT) of freshwater fish originally derived from the interspecific hybridization of red crucian carp (*Carassius auratus* red var., ♀, $2n = 100$) × common carp (*Cyprinus carpio* L., ♂, $2n = 100$). We constructed a bacterial artificial chromosome (BAC) library from allotetraploid hybrids in the 20th generation ($F_{20}$) and sequenced 14 BAC clones representing a total of 592.126 kb, identified 11 functional genes and estimated the guanine–cytosine content (37.10%) and the proportion of repetitive elements (17.46%). The analysis of intron evolution using nine orthologous genes across a number of selected fish species detected a gain of 39 introns and a loss of 30 introns in the 4nAT lineage. A comparative study based on seven functional genes among 4nAT, diploid $F_1$ hybrids ($2nF_1$) (first generation of hybrids) and their original parents revealed that both hybrid types ($2nF_1$ and 4nAT) not only inherited genomic DNA from their parents, but also demonstrated rapid genomic DNA changes (homoeologous recombination, parental DNA fragments loss and formation of novel genes). However, 4nAT presented more genomic variations compared with their parents than $2nF_1$. Interestingly, novel gene fragments were found for the *iqca1* gene in both hybrid types. This study provided a preliminary genomic characterization of allotetraploid $F_{20}$ hybrids and revealed evolutionary and functional genomic significance of allopolyploid animals.

*Heredity* (2015) **114**, 601–609; doi:10.1038/hdy.2015.3; published online 11 February 2015

## INTRODUCTION

Polyploidy, which results in organisms with multiple chromosome sets, is an important genomic event in speciation and adaptive radiation that has led to the formation of many eukaryotic lineages (Cox *et al*, 2014). Polyploids with duplicated genomes may originate from single species (autopolyploidy) or from different species through interspecific hybridization (allopolyploidy) (Otto, 2007). Allopolyploids appear prevalent in nature, suggesting an evolutionary advantage of having multiple sets of genetic material for adaptation and development (Mallet, 2007). However, much remains unknown about the processes and consequences of allopolyploidy (Abbott *et al*, 2013).

Allopolyploidy can promote the activation of cryptic mobile elements and cause rapid genomic changes (McClintock, 1984). This 'genomic shock' has been reported in many allopolyploid plants translating as gene loss, chromosome mis-pairing, retrotransposon activation, altered methylation or rearrangements between parental genomes that could lead to novel gene sequences or differential homoeologous gene expression in hybrids throughout evolution (Cox *et al*, 2014). Rapid genomic DNA changes have also been demonstrated in several allopolyploid plants. For example, homoeologous nonreciprocal recombination and biased expression of homoeologs found in allopolyploid cotton *Gossypium* sp. (Salmon *et al*, 2010), rapid genetic and epigenetic changes in allopolyploid wheat *Triticum aestivum* (Feldman and Levy, 2012) and retrotransposon activation, genomic

rearrangements and trait variation in rape *Brassica napus* (Zou *et al*, 2011). These studies provide evidence that a series of radical dynamic and stochastic genomic changes (that is, 'genomic shock') were necessary for the establishment of new allopolyploid plant species.

Although allopolyploidization is less prevalent in animals than in plants, with the development of new genetics technologies, more events of polyploidization and hybridization have been discovered and studied in animals (Otto, 2007). For example, investigating the genetic fate of duplicated *RAG* genes and rapid epigenetic changes in the allopolyploid clawed frogs *Xenopus* sp. (Evans *et al*, 2005; Chain and Evans, 2006; Koroma *et al*, 2011) or gene dosage compensation and micro RNA expression in allopolyploid *Squalius alburnoides* complex (an Iberian cyprinid fish) (Pala *et al*, 2008; Inácio *et al*, 2012). However, the manifestations of genomic shock remain less understood in the animal systems at genomic level. In addition, natural allopolyploids are usually formed hundreds or even thousands of years ago and the original diploid parental species are often extinct or difficult to identify as they continue evolving since the formation of their hybrid (Song *et al*, 1995). Thus, synthetic allopolyploids provide a model system to study and characterize radical genomic changes at early evolutionary stages.

Most genomic studies have focused on the coding sequences of functional genes, mainly because they are directly related to biological function. But non-coding regions are equally important, especially in

[1]Key Laboratory of Protein Chemistry and Fish Developmental Biology of Education Ministry of China, College of Life Sciences, Hunan Normal University, Changsha, People's Republic of China and [2]Department of Biological Sciences, National University of Singapore, Singapore, Singapore
[3]These authors contributed equally to this work.
Professor Y Liu died on 21st January 2015. This paper is dedicated to his memory.
Correspondence: Professor SJ Liu, Key Laboratory of Protein Chemistry and Fish Developmental Biology of Education Ministry of China, College of Life Science, Hunan Normal University, 36 Lushan Road, Changsha, Hunan 410081, People's Republic of China.
E-mail: lsj@hunnu.edu.cn
Received 7 August 2014; revised 17 December 2014; accepted 19 December 2014; published online 11 February 2015

the context of genome evolution (Coulombe-Huntington and Majewski, 2007). Introns possess a broad spectrum of functions and are involved in virtually every step of messenger RNA processing (Carmel and Chorev, 2012). The major objective of this research is to characterize rapid genomic changes (including intron evolution) in an allotetraploid fish hybrid lineage that was created through artificial crosses between red crucian carp (RCC) (*Carassius auratus* red var., ♀, $2n = 100$) × common carp (CC) (*Cyprinus carpio* L., ♂, $2n = 100$) and maintained for over 20 generations ($F_{20}$) (Liu *et al*, 2001). These allotetraploid fishes are currently used as the paternal progenitor to produce allotriploid fish for aquaculture purposes. Our previous studies demonstrated that $F_1$ and $F_2$ progenies were fertile diploid hybrids ($2n = 100$). However, from $F_3$ onwards, fertile allotetraploid individuals of both sexes were produced (Liu *et al*, 2001; Liu, 2010). This allotetraploid hybrid fish system is therefore unique and provides an opportunity to test which genetic elements are susceptible to rapid genomic changes. Our previous studies also demonstrated the loss of paternal DNA fragments and a recombined mitochondrial DNA segment in this allopolyploid hybrid lineage (Liu, 2010). In the present study, we performed a comparative study on this artificially derived allotetraploid lineage (4nAT) with parental species and hybrids of first generation ($F_1$), with the expectation to provide a new perspective on genomic evolution in allopolyploid animals.

## MATERIALS AND METHODS
### BAC library construction and sequencing
DNA samples of five individuals from the 20th generation progeny of an allotetraploid hybrid lineage (4nAT) were used to construct a bacterial artificial chromosome (BAC) library. The fish were sampled from Engineering Research Center of Polyploid Fish Breeding at Hunan Normal University (Changsha, China). A total of 10 ml fresh peripheral blood from each fish was used for DNA extraction using the method previously reported to generate high-molecular weight genomic DNA (Katagiri *et al*, 2000). The homogenized blood sample was mixed with an equal volume of pre-warmed 1% low melting point agarose at a concentration of $5 \times 10^8$ cells ml$^{-1}$, and cast into plugs using plug molds (Bio-Rad, Guangzhou, China). The agarose plugs were cut and digested with the restriction enzyme *Hind*III, DNA fragments (100–300 kb) were ligated to CopyControl pCC1BAC (*Hind*III Cloning-Ready) Vector (Epicentre, Madison, WI, USA), and then transformed to *Escherichia coli DH10B*. All positive BAC clones were hand-picked and stored at −80 °C. A total of 14 BAC clones were randomly chosen for shotgun sequencing. Subclone libraries were constructed first with insert fragment size 1–2 kb and sequenced on an ABI3730xl platform by Majorbio Bio-pharm Technology Co. Ltd (Shanghai, China). The computer programs PHRED, PHRAP and CONSED were used to perform base calling and quality assessment, sequence assembly and contigs ordering for each BAC clone (Ewing and Green, 1998; Gordon *et al*, 1998). The programs were used under their default settings.

### Characterization of BAC clone sequences
Guanine–cytosine content of each BAC clone sequence was calculated using a custom Perl script. Repetitive DNA content was estimated using RepeatMasker (version 4.0.5, http://www.repeatmasker.org) against the REPBASE repeat database (20140131). The program FGENESH (Salamov and Solovyev, 2000) was used to predict gene structure of the 14 BAC sequences using zebrafish *Danio rerio* as reference organism. Predicted genes were annotated using BLASTX against the NCBI-NR (National Center for Biotechnology Information-Non-redundant) protein database with an *E*-value < 1E-10.

### Evolution of intron density
The intron information (for example, intron number, intron length) was extracted from the gene structure predicted for each BAC sequence, and intron density was calculated based on number of introns per gene. To infer the evolutionary trend of intron in 4nAT across the fish lineage, comparative intron analysis was conducted among 4nAT and other fish species, by selecting annotated orthologous sequences from zebrafish (DRE), medaka (*Oryzias latipes*), fugu (*Takifugu rubripes*), tetraodon (*Tetraodon nigroviridis*), tilapia (*Oreochromis niloticus*), sea lamprey (*Petromyzon marinus*, PMA), three-spined stickleback (*Gasterosteus aculeatus*), Atlantic cod (*Gadus morhua*), Human (*Homo sapiens*) and western clawed frog (*Xenopus tropicalis*), and used *H. sapiens* and *X. tropicalis* as outgroups. Orthologous gene sequences and detailed intron information were downloaded from Ensembl website (http://asia.ensembl.org/index.html). Multiple alignments of protein sequences with known intron locations and lengths were used to analyze intron evolution using Malin software (http://www.iro.umontreal.ca/~csuros/introns/malin/). The guide tree was constructed based upon *ADP-ribosylation factor-like 14* gene sequences. The Dollo parsimony method was used to estimate intron loss and gain events (Csűrös, 2008).

### Homologous gene amplification in parental and hybrid fishes
To verify the accuracy of sequencing assembly and estimate the genomic changes, all predicted functional genes were amplified from 4nAT, diploid $F_1$ hybrids ($2nF_1$), RCC and CC. Polymerase chain reaction primers were designed based on the conserved regions of DRE, *O. latipes*, *T. rubripes*, *T. nigroviridis*, *O. niloticus* and 4nAT across different exons (Supplementary Table S1). Genomic DNA of 4nAT, $2nF_1$, RCC and CC were extracted from peripheral blood using the Ezup Column Blood Genomic DNA Extraction Kit (Sangon Biotech, Shanghai, China). Long polymerase chain reaction was carried out using TaKaRa PrimeSTAR GXL DNA Polymerase (TaKaRa, Dalian, China), with the following polymerase chain reaction settings: 94 °C for 1 min; 30 cycles of 98 °C for 10 s, 68 °C for *N* min (where *N* represents the annealing time for each gene) and 72 °C for 10 min. Polymerase chain reaction products within the expected size range were extracted and purified using SanPrep Column DNA Gel Extraction Kit (Sangon Biotech). The purified DNA fragments were inserted into vector PMD18-T (TaKaRa) and transformed into DH5α competent cells. A total of 10 clones for each gene insert in 4nAT, $2nF_1$, RCC and CC were sequenced on the ABI3730 sequencer (GenScript Corporation, Nanjing, China). Gene sequences were named with the abbreviated fish name followed by Roman numbers. For example, if RCC had two gene copies, they would be named RCC-I and RCC-II, respectively.

### Sequence comparison and analysis
Sequences of complementary DNA for each gene were verified and annotated through Basic Local Alignment Search Tool X (BLASTX) search against NCBI-NR database (http://blast.ncbi.nlm.nih.gov) and Ensembl zebrafish DRE protein database (http://uswest.ensembl.org/Tools/Blast?db=core). Multiple sequence alignment was performed to assess sequence similarity and variation among parental species (RCC, CC) and hybrids (4nAT, $2nF_1$) using ClustalW (Thompson *et al*, 1994). MEGA 5 was used to construct phylogenetic trees based on exon, intron and whole-DNA sequences using the maximum likelihood method with 1000 bootstraps (Tamura *et al*, 2011). Protein structures were predicted through the SWISS-MODEL web server (Arnold *et al*, 2006). To identify homoeologous recombination in hybrids, diagnostic single-nucleotide polymorphisms (SNPs) for RCC and CC were first identified, and then SNP biases were examined in the sequences of hybrids. In addition, we defined two genetic patterns to assess parental origin of hybrid sequences: (i) 'genetic inheritance' in case hybrid gene sequences were identical to either parental species; and (ii) 'genetic variation' in case large sequence variation was found between hybrids and their parental species, such as homoeologous recombination, DNA fragments insertion and deletion.

## RESULTS
### Characterization of BAC clone sequences
A selection of 384–576 subclones with fragment inserts of 1–2 kb were sequenced and assembled for each of the 14 BAC clone sequences (NCBI accession numbers: KF758440-KF758444 and KJ424354-KJ424362). The lengths of the assembled BAC clone sequences ranged from 17 849 bp to 87 725 bp, with an average of 42 295 bp. Guanine–cytosine content of the BAC clones ranged from 34.38 to 39.80% with an average of 37.10% (Table 1).

From the 14 assembled BAC clone sequences, 103 411 bp out of the total 592 126 bp were identified as repetitive sequences (17.46%). The classification and corresponding proportion of different repetitive

elements were shown in Table 2. The most abundant type of repetitive elements in allotetraploid hybrids (4nAT) was retroelements (6.47%) followed by DNA transposons (4.17%). Other repeats include 345 simple repeats (3.67%), 118 low-complexity repeats (0.85%) and 105 small RNA repeat (1.59%). In comparison with CC, the only parental species for which a BAC library has been constructed and analyzed (Xu *et al*, 2011), 4nAT hybrids presented a higher proportion of retroelements and a lower proportion of DNA transposons (Table 2).

Eleven functional genes were predicted (Table 3) with total lengths ranging from 609 bp to 41 004 bp, and coding DNA sequence regions of 609 bp to 3579 bp long. Two genes (*ADP-ribosylation factor-like* 14 and *chemokine (C-X-C motif) receptor 7b*) had no intron. The *importin-13-like* gene evidenced the largest number of exons and introns (28 exons and 27 introns). The amino-acid sequence similarity between 4nAT and zebrafish DRE ranged from 43 to 96%, depending on specific genes (Table 3).

### Evolution of intron density

Intron information of orthologous genes in different species was detailed in Table 4. The total number of introns ranged from 81 to 112, and the intron density was from 9.4 in 4nAT to 12.9 in *G. aculeatus*, respectively. The average intron length in 4nAT (8564 bp) was larger than in the selected fish species except for DRE, *G. morhua* and PMA. The analysis of intron evolution demonstrated that the introns of the nine 4nAT genes evolved dynamically with a net gain of 30 introns and a loss of 39 introns. When compared to PMA or to

**Table 1 Sequence information of allotetraploid hybrids' (4nAT) BAC clones**

| BAC clone ID | BAC clone size (bp) | GC content (%) | GenBank accession number |
|---|---|---|---|
| 17-B4 | 31,067 | 37.03 | KF758440 |
| 17-E7 | 57,220 | 36.56 | KF758441 |
| 136-A3 | 44,360 | 39.80 | KF758442 |
| 136-B4 | 45,733 | 35.75 | KF758443 |
| 136-E7 | 25,382 | 37.85 | KF758444 |
| 149-B4 | 25,308 | 36.54 | KJ424354 |
| 149-B8 | 45,494 | 34.38 | KJ424355 |
| 149-D6 | 18,748 | 39.35 | KJ424356 |
| 149-F8 | 39,387 | 36.36 | KJ424357 |
| 150-B4 | 87,725 | 37.03 | KJ424358 |
| 150-B8 | 34,932 | 36.21 | KJ424359 |
| 150-D6 | 42,527 | 39.03 | KJ424360 |
| 150-F8.1 | 76,394 | 35.03 | KJ424361 |
| 150-F8.2 | 17,849 | 38.46 | KJ424362 |

Abbreviations: BAC, bacterial artificial chromosome; GC, guanine–cytosine.

**Table 2 Repetitive DNA elements detected in allotetraploid hybrids' (4nAT) BAC clone sequences**

| Type | Subtype | Number of repeats | Total length (bp) | Proportion of repeats (%) | Proportion of repeats in common carp (%)[a] |
|---|---|---|---|---|---|
| Retroelements | | 67 | 38,306 | 6.47% | 4.52% |
| | LTR | 31 | 21,690 | 3.66% | 1.98% |
| | LINEs | 35 | 16,106 | 2.72% | 2.33% |
| | SINEs | 1 | 510 | 0.09% | 0.20% |
| DNA transposons | | 199 | 24,691 | 4.17% | 6.67% |
| Simple repeats | | 345 | 21,747 | 3.67% | 1.64% |
| Small RNA | | 105 | 9422 | 1.59% | 0.11% |
| Satellites | | 48 | 5957 | 1.01% | 2.46% |
| Low complexity | | 118 | 5011 | 0.85% | 1.98% |
| Unclassified | | 19 | 1856 | 0.31% | 0.05% |

Abbreviations: LINE, long interspersed nuclear element; LTR, long terminal repeat; SINE, short interspersed nuclear element.
[a]Xu *et al*, 2011.

**Table 3 Gene annotation of allotetraploid hybrids' (4nAT) BAC clone sequences and comparison with zebrafish *Danio rerio* orthologous sequences**

| Gene ID | Gene length (bp) | CDS length (bp) | Number of exons (4nAT/zebrafish) | Gene name | Sequence similarity (amino acid) |
|---|---|---|---|---|---|
| 1 | 13,235 | 2658 | 13/20 | Integrator complex subunit 7 | 84% |
| 2 | 5719 | 1929 | 15/15 | Denticleless homolog | 82% |
| 3 | 7543 | 1380 | 13/14 | Fizzy-related protein homolog | 96% |
| 4 | 17,416 | 3579 | 28/22 | Importin-13-like | 48% |
| 5 | 1258 | 1011 | 2/3 | Guanine nucleotide-binding protein G(I)/G(S)/G(O) subunit gamma-5 | 94% |
| 6 | 609 | 609 | 1/1 | ADP-ribosylation factor-like 14 | 88% |
| 7 | 2971 | 948 | 6/8 | Iqca1 | 43% |
| 8 | 1089 | 1089 | 1/1 | Chemokine (C-X-C motif) receptor 7b | 91% |
| 9 | 41,004 | 1404 | 15/19 | Phosphodiesterase 11 A | 76% |
| 10 | 1550 | 1056 | 2/3 | Uncharacterized protein LOC101883163 | 91% |
| 11 | 7192 | 2382 | 17/18 | wu:fk36e11 | 83% |

**Table 4** Intron information of orthologous genes identified in allotetraploid hybrids (4nAT) and other fish species

| Taxa | Number of genes[a] | Number of introns | Intron density (number of introns/gene) | Average intron length (bp) |
|---|---|---|---|---|
| *Danio rerio* | 9 | 94 | 10.4 | 32,164 |
| Allotetraploid hybrid lineage | 9 | 85 | 9.4 | 8564 |
| *Oryzias latipes* | 7 | 81 | 11.5 | 4749 |
| *Gadus morhua* | 9 | 112 | 12.4 | 11,189 |
| *Petromyzon marinus* | 7 | 112 | 16 | 26,107 |
| *Oreochromis niloticus* | 9 | 104 | 11.5 | 8175 |
| *Takifugu rubripes* | 9 | 95 | 10.6 | 4948 |
| *Tetraodon nigroviridis* | 9 | 108 | 12 | 2971 |
| *Gasterosteus aculeatus* | 8 | 103 | 12.9 | 4814 |

[a]Not all species have nine orthologous genes.
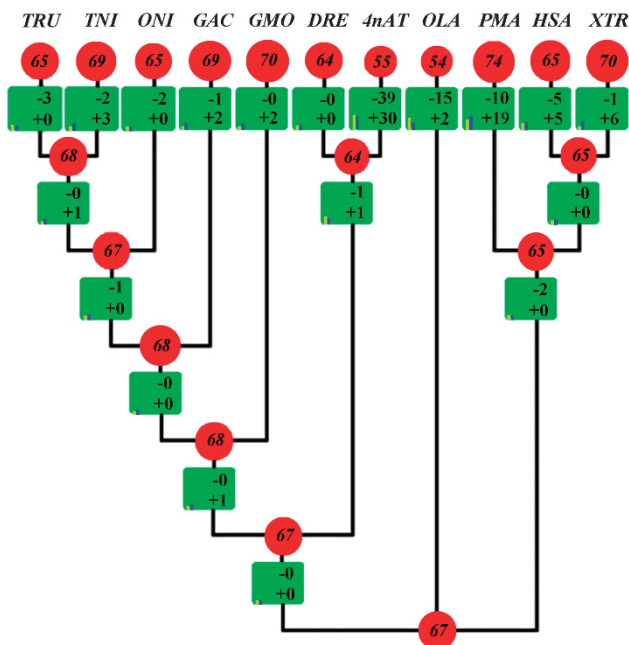


**Figure 1** Dollo parsimony prediction of intron densities for orthologous genes among fish species. Numbers in red circles indicate conserved intron numbers in a species or a phylogenetic clade, with the area of red circles proportional to the corresponding numbers; numbers with ' – ' and '+' in green boxes indicate number of introns lost and gained, respectively. TRU: *Takifugu rubripes*, TNI: *Tetraodon nigroviridis*, ONI: *Oreochromis niloticus*, GAC: *Gasterosteus aculeatus*, GMO: *Gadus morhua*, DRE: *Danio rerio*, 4nAT: allotetraploid hybrid lineage, OLA: *Oryzias latipes*, PMA: *Petromyzon marinus*, HSA: *Homo sapiens* and XTR: *Xenopus tropicalis*.

O. latipes genes, a net gain of 19 and a loss of 10 introns and a net gain of 2 and a loss of 15 introns were found in the two fish species, respectively (Figure 1). The number of introns shared between 4nAT and the other fish species ranged from 14 to 25. However, when comparing DRE with the remaining fish species, this value ranged from 49 to 63 introns (Figure 2), which suggests rapid intron evolution in the 4nAT.

## Comparative analysis of functional genes among parental species and hybrids

The sequences of seven homologous genes for 4nAT, 2nF$_1$, RCC and CC are available at NCBI GenBank (Accession numbers: KF769270-KF769301 and KM088001-KM088012). Different copies of these genes were described in all the four species/lineages (Table 5). Eight models
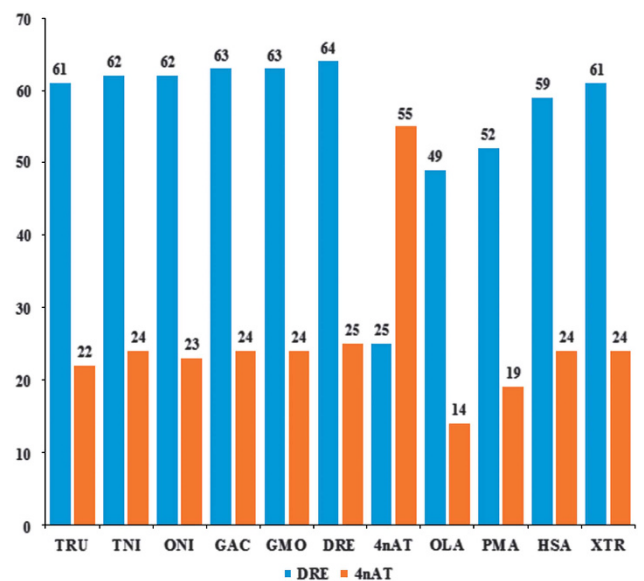


**Figure 2** Number of shared introns between model fish species and allotetraploid hybrids (4nAT) (orange) or *Danio rerio* (blue) predicted by the Malin software based on the nine orthologous genes. TRU: *Takifugu rubripes*, TNI: *Tetraodon nigroviridis*, ONI: *Oreochromis niloticus*, GAC: *Gasterosteus aculeatus*, GMO: *Gadus morhua*, 4nAT: allotetraploid hybrid lineage, OLA: *Oryzias latipes*, PMA: *Petromyzon marinus*, HSA: *Homo sapiens* and XTR: *Xenopus tropicalis*.

of genetic change could be associated with hybrid sequence inheritance, with models 1 and 2 belonging to the 'genetic inheritance' pattern and models 3–8 attributed to the 'genetic variation' pattern (Figure 3). Four genes of 2nF$_1$ and two genes of 4nAT were associated with the 'genetic inheritance' pattern. Five genes showed the 'genetic variation' pattern in at least one sequence copy in 4nAT, whereas only three genes of 2nF$_1$ demonstrated rapid sequence change in 4nAT.

Two genes (*fizzy-related protein homolog* and *importin-13-like*) fell strictly within the 'genetic inheritance' pattern in both 4nAT and 2nF$_1$ hybrids. The 2nF1-II and 4nAT copies of the *fizzy-related protein homolog* gene showed a higher similarity to CC (97.2% and 97.0%, respectively) (Figure 3—Model 2, Supplementary Figure S8A), whereas 2nF1-I showed a relatively higher sequence similarity to CC (73.8%) than to RCC (55.8%) (Supplementary Table S2, Supplementary Figure S1). The 2nF1-I copy of the *importin-13-like* gene was found to have been inherited from RCC (99.8%), whereas the 2nF1-II copy was highly similar to CC-II (98.0%). On the other

**Table 5** Information on the homologous genes from allotetraploid hybrids (4nAT), diploid hybrids (2nF$_1$), red crucian carp (RCC) and common carp (CC)

| Gene name | Exons amplified | Taxa | Sequence name | Sequence length (bp) | Inheritance pattern[a] | GenBank accession number |
|---|---|---|---|---|---|---|
| Denticleless homolog | Exon3–exon15 | RCC | RCC-I | 5371 | NA | KF769276 |
| | | | RCC-II | 5239 | NA | KF769277 |
| | | CC | CC | 5128 | NA | KF769278 |
| | | 2nF$_1$ | 2nF1 | 4928 | Genetic variation | KM088003 |
| | | 4nAT | 4nAT-I | 5301 | Genetic variation | KF769279 |
| | | | 4nAT-II | 5418 | Genetic variation | KF769280 |
| Guanine nucleotide-binding protein G(I)/G(S)/G(O) subunit gamma-5 | Exon1–exon2 | RCC | RCC | 662 | NA | KF769284 |
| | | CC | CC | 657 | NA | KF769285 |
| | | 2nF$_1$ | 2nF1-I | 665 | Genetic inheritance | KM088006 |
| | | | 2nF1-II | 657 | Genetic inheritance | KM088007 |
| | | 4nAT | 4nAT-I | 655 | Genetic inheritance | KF769286 |
| | | | 4nAT-II | 662 | Genetic inheritance | KF769287 |
| | | | 4nAT-III | 662 | Genetic variation | KF769288 |
| Fizzy-related protein homolog | Exon1–exon11 | RCC | RCC | 5617 | NA | KF769281 |
| | | CC | CC | 5348 | NA | KF769282 |
| | | 2nF$_1$ | 2nF1-I | 5413 | Genetic inheritance | KM088004 |
| | | | 2nF1-II | 5370 | Genetic inheritance | KM088005 |
| | | 4nAT | 4nAT | 5326 | Genetic inheritance | KF769283 |
| Chemokine (C-X-C motif) receptor 7b | No introns | RCC | RCC-I | 958 | NA | KF769270 |
| | | | RCC-II | 959 | NA | KF769271 |
| | | CC | CC | 959 | NA | KF769272 |
| | | 2nF$_1$ | 2nF1-I | 959 | Genetic variation | KM088001 |
| | | | 2nF1-II | 959 | Genetic inheritance | KM088002 |
| | | 4nAT | 4nAT-I | 959 | Genetic inheritance | KF769273 |
| | | | 4nAT-II | 959 | Genetic inheritance | KF769274 |
| | | | 4nAT-III | 959 | Genetic variation | KF769275 |
| Iqca1 | Exon2–exon 4 | RCC | RCC | 988 | NA | KF769293 |
| | | CC | CC | 992 | NA | KF769294 |
| | | 2nF$_1$ | 2nF1 | 789 | Genetic variation | KM088008 |
| | | 4nAT | 4nAT | 789 | Genetic variation | KF769295 |
| Importin-13-like | Exon10–exon13 | RCC | RCC | 1716 | NA | KF769289 |
| | | CC | CC-I | 1735 | NA | KF769290 |
| | | | CC-II | 1800 | NA | KF769291 |
| | | 2nF$_1$ | 2nF1-I | 1715 | Genetic inheritance | KM088009 |
| | | | 2nF1-II | 1795 | Genetic inheritance | KM088010 |
| | | 4nAT | 4nAT | 1711 | Genetic inheritance | KF769292 |
| Phosphodiesterase 11A | Exon8–exon15 | RCC | RCC-I | 5121 | NA | KF769296 |
| | | | RCC-II | 5093 | NA | KF769297 |
| | | CC | CC-I | 5872 | NA | KF769298 |
| | | | CC-II | 5980 | NA | KF769299 |
| | | 2nF$_1$ | 2nF1-I | 5145 | Genetic inheritance | KM088011 |
| | | | 2nF1-II | 5195 | Genetic inheritance | KM088012 |
| | | 4nAT | 4nAT-I | 5674 | Genetic variation | KF769300 |
| | | | 4nAT-II | 5511 | Genetic variation | KF769301 |

Abbreviations: CC, common carp; RCC, red crucian carp; NA, not applicable.
[a]Genetic inheritance: sequences identical between hybrids and parental species; genetic variation: large variation between hybrids and parental sequences including parental DNA fragments loss, homoeologous recombination and the formation of novel genes.

hand, sequence similarity of *importin-13-like* gene was much higher between 4nAT and RCC (99.6%) than with CC-I (67.9%) or CC-II (70.8%) (Supplementary Table S3, Supplementary Figure S2, Figure 3 —Model 1, Supplementary Figure S8B).

Two genes (*denticleless homolog* and *iqca1*) fell strictly within the 'genetic variation' pattern in both 4nAT and 2nF$_1$ hybrids. The alignment and phylogenetic analyses of the *denticleless homolog* gene (Supplementary Figure S3) demonstrated that both 4nAT-I and 4nAT-II sequences underwent homoeologous recombination

(Figure 3—Model 5 and 6). Results based on exon sequences clustered 4nAT-II with CC, however, using intron and whole-gene sequences, 4nAT was grouped with RCC (Supplementary Figure S8C). This discrepancy is indicative of homoeologous recombination also detected in the *denticleless homolog* gene copy of 2nF$_1$ hybrids but with more mutations when compared with RCC and CC sequences (Supplementary Figure S3, Supplementary Table S4). The *iqca1* gene had a single copy in all four fish species/lineages, with many mutations, deletions and insertions found in 2nF$_1$ and 4nAT hybrids
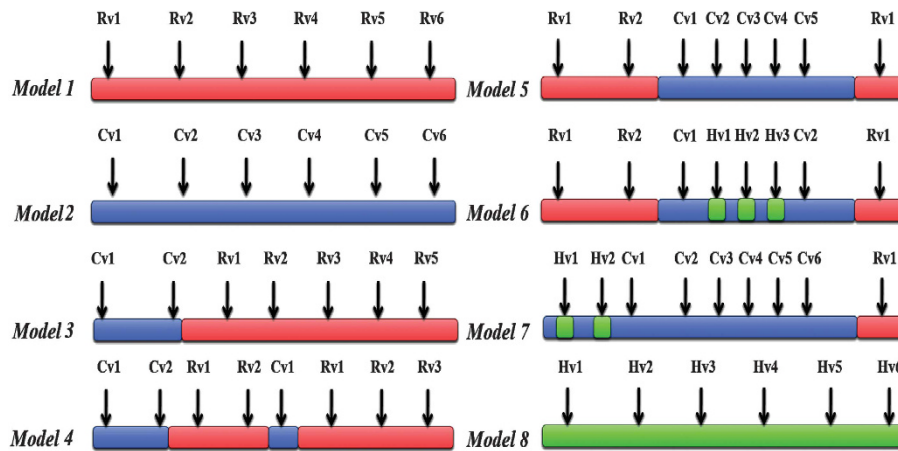
**Figure 3** Patterns of genetic inheritance (Model 1 and 2) and genetic variation (Model 3–8) hypothesized for allotetraploid hybrids (4nAT). The red color indicates homoeologous sequences inherited from red crucian carp (RCC), whereas the blue color indicates homoeologous sequences inherited from common carp (CC). The green color indicates novel sequences in 4nAT. Rv, Cv, and Hv denote SNP variation identical to RCC, CC or 4nAT, respectively. Numbers following Rv, Cv and Hv indicate hypothetical numbers of SNP variation.

(Figure 3—Model 8, Supplementary Table S5,Supplementary Figure S4 and Supplementary Figure S8D). Particularly, extensive deletions in exon 2 and intron 2 (Figure 4) and altered protein structure (4nAT and $2nF_1$) (Supplementary Figure S9) indicate marked sequence change as a consequence of hybridization.

As regard to the *chemokine (C-X-C motif) receptor 7b* gene, both 'genetic inheritance' and 'genetic variation' patterns were implicated in $2nF_1$ and 4nAT hybrid lineages. Sequence similarity was very high between 2nF1-II and CC (99.0%), 4nAT-I and RCC-II (99.2%) and between 4nAT-II and CC (96.8%) (Supplementary Table S6, Supplementary Figure S5). Although 2nF1-I and 4nAT-III showed higher sequence similarity to RCC-I (96.4% and 96.3%, respectively), fragments of homoeologous recombination with CC were also detected (strong SNP bias at positions 1–150 bp and 459–542 bp) (Figure 3—Model 4), resulting in three phylogenetic clusters: (i) 2nF1-II, CC and 4nAT-II; (ii) 4nAT-I and RCC-II; and (iii) 2nF1-I, 4nAT-III and RCC-I (Figure 5).

The *guanine nucleotide-binding protein G(I)/G(S)/G(O) subunit gamma-5* gene and *phosphodiesterase 11 A* gene showed a strict pattern of 'genetic inheritance' in $2nF_1$ hybrids, whereas in 4nAT, both patterns were found for the first gene and only 'genetic variation' for the second gene. Multiple sequence alignment analysis of *guanine nucleotide-binding protein G(I)/G(S)/G(O) subunit gamma-5* gene (Supplementary Figure S6) revealed higher similarity between 2nF1-I and RCC (98.6%), between 2nF1-II and CC (98.9%) (Supplementary Table S7), between 4nAT-I and CC (99.3%) and between 4nAT-II and RCC (99.6%) (Supplementary Table S7, Supplementary Figure S8E). Homoeologous recombination was detected in 4nAT-III with a SNP bias towards CC from 1 to 60 bp (Supplementary Figure S6) and the remaining SNPs biased towards RCC (Figure 3—Model 3). Homoeologous recombination was also found in two sequences of the 4nAT *phosphodiesterase 11 A* gene (Supplementary Figure S7): SNPs from 1-4366 bp of 4nAT-I were similar to CC-I, whereas SNPs starting from 4367 bp were similar to RCC-II (Figure 3—Model 7); from 1 to 1991 bp of 4nAT-II, the sequence was similar to RCC-II with only few mutations, from 1992 to 4365 bp it was similar to CC and again from 4366 bp onwards it was similar to RCC-II (Figure 3—Model 5). 2nF1-I and 2nF1-II both showed higher sequence similarity with RCC-I (93.6% and 90.8%, respectively) (Supplementary Table S8, Supplementary Figure S8F).
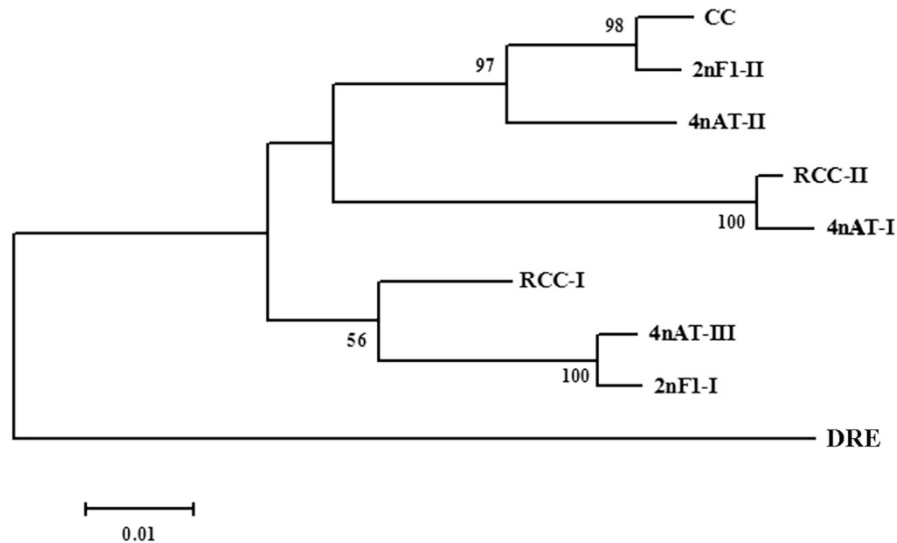
## DISCUSSION

### Genomic properties of 4nAT BAC clone sequences

Allotetraploid hybrids are characterized by doubled genomes. The increased genome size could produce duplicated functional genes that can also be accompanied by rapid and extensive genomic DNA changes and gene silencing (Chain and Evans, 2006). DNA methylation, a critical mechanism of gene silencing, usually occurs in promoter regions as a means of regulating the expression level of the duplicated genes for them to stabilize in the allotetraploid population and promote evolution (Sehrish et al, 2014). This process might explain variant densities and distributions of CpG islands among fish genomes, which can cause guanine–cytosine content variation (Han and Zhao, 2008). In the clawed frogs *Xenopus* sp. hybrid system, a higher proportion of methylated fragments were found in the hybrids compared with the parental species (Koroma et al, 2011). Higher guanine–cytosine content was also found in the allotetraploid hybrid 4nAT in contrast to its male progenitor CC (Xu et al, 2011) that could be associated with an elevated level of DNA methylation levels (Xiao et al, 2013). However, given that only a small portion of the genome was surveyed, the definite assentation that DNA methylation is more prevalent in the allotetraploid hybrids versus parental species requires further analysis of other genomic regions.

The proportion of all repetitive elements was comparable between 4nAT and CC (Xu et al, 2011) but 4nAT were characterized by a higher proportion of retroelements (Table 3). The duplication and/or transposition of retroelements into new sites may directly affect gene structure, and the presence of multiple copies of these elements throughout the genome could have long-term effects on recombination events and a more subtle influence on gene expression (Feschotte, 2008). The insertion of retro-elements within genes could cause gene inactivation through disruption of the reading frame or promoter regions, another mechanism to turn off one copy of duplicated genes in allopoly-ploid species (Casacuberta and González, 2013). In the sunflower *Helianthus* sp. the genome size is at least 50% larger in hybrids than in pure species owing to retrotransposon proliferation (Ungerer et al, 2006). The genome size of allotetraploid hybrids (4nAT) (*C*-value: ~ 3.86 pg) is estimated to be larger than RCC (*C*-value: 1.88–2.14 pg) and CC (*C*-value: 1.61–2.03 pg)

*Exon 2 alignment*



*Intron 2 alignment*



**Figure 4** Sequence alignment of *iqca1* exon 2 and intron 2 from red crucian carp (RCC), common carp (CC), diploid F$_1$ hybrids (2nF$_1$) and allotetraploid hybrids (4nAT). Red background indicates deletion, black background with dots indicates identical region.

(http://www.genomesize.com) which could be associated with the proportion of retroelements (Table 3). Our limited sequence data at the moment, however, does not allow us to make any further assertions.

**Evolution of intron density**

Comparative intron evolution analysis was conducted among 4nAT and other fish species in this study to assess the rate of intron evolution in 4nAT. Our results suggested that the smallest intron

Figure 5 Genetic relationship among *chemokine (C-X-C motif) receptor 7b* sequences from red crucian carp (RCC), common carp (CC), diploid $F_1$ hybrids ($2nF_1$) and allotetraploid hybrids (4nAT) using *Danio rerio* (DRE) as outgroup. Bootstrap values above 50 and percent nucleotide substitution were shown.

density found in 4nAT compared with other fish species could be attributed to DNA loss, a common mechanism of genome stabilization in allopolyploid hybrids (Buggs *et al*, 2012). The relatively larger intron length found in 4nAT hybrids in comparison with selected fish species except for DRE, *G. morhua* and PMA can be explained by doubled genome and larger genome size (*C*-value: ~ 3.86 pg) of 4nAT. On the other hand, the largest intron density found in PMA may relate to the preservation of ancestral introns (Kawaguchi *et al*, 2010) and the larger average intron length in DRE is likely to be mainly due to intron size expansion events (Moss *et al*, 2011).

A previous study on intron evolution in animals attributed mammals' differences in gene structure to intron loss but not to intron gain (Coulombe-Huntington and Majewski, 2007). Our results revealed several intron gain events especially in 4nAT and PMA. Being a primitive vertebrate species and the ancestor of all teleost fishes it is expected of PMA to possess ancient introns that might have been lost in most divergent fish (Venkatesh *et al*, 1999). The dynamics of intron gain and loss in 4nAT (39 gained and 30 lost introns) showed a rapid genomic change in these hybrids. A number of factors including genome size, breeding cycle, gene expression level and intron length are related to intron gain and loss events (Coulombe-Huntington and Majewski, 2007). However, interspecific hybridization and polyploidization are likely the most crucial factors in the rapid intron evolution observed in 4nAT possibly allowing for the coordinated expression of duplicated genes (Carmel and Chorev, 2012).

### Models of rapid genomic change

Three genetic models of rapid genomic changes were identified in 4nAT hybrids including the loss of parental DNA fragments, homoeologous recombination and the formation of novel genes. The loss of parental DNA fragments has been commonly observed in allopolyploid plants (Buggs *et al*, 2012; Sehrish *et al*, 2014). Similarly, it seems to have happened during the establishment of the 4nAT allotetraploid hybrid lineage, as only one copy of *fizzy-related protein homolog* gene and one copy of *importin-13-like* gene were found even though two copies for each gene were present in the $2nF_1$. Previous inter simple sequence repeat and amplified fragment length polymorphism studies indicated a bias of DNA loss towards the paternal (CC) genome (Liu, 2010). This observation agrees with the findings in triticale (a hybrid

of wheat and rye) (Ma and Gustafson, 2006) and cordgrass *Spartina* sp. (Salmon *et al*, 2005). In this study only two genes evidenced DNA fragments loss, one from each parent, thus hindering the evaluation of potential parental bias. Similarly, research on clawed frogs *Xenopus* sp. showed no evidence of directional loss of sequences towards either parental species (Koroma *et al*, 2011). To understand if this might be a particularity of allopolyploid animals, further larger scale genomic studies are required to address this issue.

Genetic recombination is a process that generates novelty contributing to genetic variation and genome structural diversity in organisms (Gaut *et al*, 2007). In allopolyploids, recombination usually occurs among paralogous or homoeologous sequences and it has been shown as a major cause for genomic changes such as DNA deletion, duplication and gene conversion (Gaeta and Chris Pires, 2010). Genetic recombination has been reported in both allopolyploid plants such as rape *Brassica napus* (Zou *et al*, 2011) and cotton *Gossypium* sp. (Salmon *et al*, 2010) and allopolyploid animals, for example, clawed frogs and salamanders (Evans *et al*, 2005; Bi *et al*, 2008). In the present study, many variations, such as DNA deletions and insertions in 4nAT-I sequences of *phosphodiesterase 11 A* and *denticleless homolog* genes are suspected to be the consequence of homoeologous recombination (Figure 3—Model 6). Homoeologous recombination was detected in four out of seven genes in allotetraploid hybrids (4nAT) versus only one out of seven in $2nF_1$, strongly suggesting this as a potentially important mechanism for the rapid genomic changes observed in 4nAT.

Novel fragments were commonly observed in allopolyploid plants, which have been considered a critical mechanism for adaption and evolution after genome duplication (Chen, 2007). The novel *iqca1* gene identified in both $2nF_1$ and 4nAT hybrids when compared with parental species RCC and CC seems to have resulted from DNA deletions, insertions and mutations (Supplementary Figure S4, Figure 4). The Iqca1 protein structure was different in both hybrids in comparison with the parental species (Supplementary Figure S9), confirming a novel gene function in the hybrids. This, in turn, provides additional evidence to support that interspecific hybridization could foster large genomic changes.

### CONCLUSION

Few allopolyploidization studies focused on animals. We used an artificially derived allotetraploid lineage of freshwater fish to help filling

this gap. We sequenced and assembled 14 BAC clones of $F_{20}$ allotetraploid hybrids (4nAT), analyzed the evolution of introns and compared seven genes across the parental species CC and RCC as well as on $2nF_1$. This study demonstrated that rapid genomic changes are facilitated by intron gain and loss, homoeologous recombination and the formation of novel genes. Large-scale genomic studies are needed to verify these findings. Nevertheless, this study provided a preliminary genomic characterization of allotetraploid $F_{20}$ hybrids, revealing evolutionary and functional genomic significance of allopolyploid animals.

## DATA ARCHIVING

Sequencing data from this article have been deposited in GenBank under the accession numbers: KF758440 to KF758444, KJ424354 to KJ424362, KF769270 to KF769301 and KM088001 to KM088012.

## CONFLICT OF INTEREST

The authors declare no conflict of interest.

Abbott R, Albach D, Ansell S, Arntzen JW, Baird SJE, Bierne N et al. (2013). Hybridization and speciation. J Evol Biol 26: 229–246.

Arnold K, Bordoli L, Kopp J, Schwede T (2006). The SWISS-MODEL workspace: a web-based environment for protein structure homology modelling. Bioinformatics 22: 195–201.

Bi K, Bogart J, Fu J (2008). The prevalence of genome replacement in unisexual salamanders of the genus Ambystoma (Amphibia, Caudata) revealed by nuclear gene genealogy. BMC Evol Biol 8: 158.

Buggs RJA, Chamala S, Wu W, Tate JA, Schnable PS, Soltis DE et al. (2012). Rapid, repeated, and clustered loss of duplicate genes in allopolyploid plant populations of independent origin. Curr Biol 22: 248–252.

Carmel L, Chorev M (2012). The function of introns. Front Genet 3: 55.

Casacuberta E, González J (2013). The impact of transposable elements in environmental adaptation. Mol Ecol 22: 1503–1517.

Chain FJJ, Evans BJ (2006). Multiple mechanisms promote the retained expression of gene duplicates in the tetraploid frog Xenopus laevis. PLoS Genet 2: e56.

Chen ZJ (2007). Genetic and epigenetic mechanisms for gene expression and phenotypic variation in plant polyploids. Annu Rev Plant Biol 58: 377–406.

Coulombe-Huntington J, Majewski J (2007). Characterization of intron loss events in mammals. Genome Res 17: 23–32.

Cox MP, Dong T, Shen G, Dalvi Y, Scott DB, Ganley ARD (2014). An interspecific fungal hybrid reveals cross-kingdom rules for allopolyploid gene expression patterns. PLoS Genet 10: e1004180.

Csűrös M (2008). Malin: maximum likelihood analysis of intron evolution in eukaryotes. Bioinformatics 24: 1538–1539.

Evans BJ, Kelley DB, Melnick DJ, Cannatella DC (2005). Evolution of RAG-1 in polyploid clawed frogs. Mol Biol Evol 22: 1193–1207.

Ewing B, Green P (1998). Base-calling of automated sequencer traces using phred.II. Error probabilities. Genome Res 8: 186–194.

Feldman M, Levy AA (2012). Genome evolution due to allopolyploidization in wheat. Genetics 192: 763–774.

Feschotte C (2008). Transposable elements and the evolution of regulatory networks. Nat Rev Genet 9: 397–405.

Gaeta RT, Chris Pires J (2010). Homoeologous recombination in allopolyploids: the polyploid ratchet. New Phytol 186: 18–28.

Gaut BS, Wright SI, Rizzon C, Dvorak J, Anderson LK (2007). Recombination: an underappreciated factor in the evolution of plant genomes. Nat Rev Genet 8: 77–84.

Gordon D, Abajian C, Green P (1998). Consed: a graphical tool for sequence finishing. Genome Res 8: 195–202.

Han L, Zhao Z (2008). Comparative analysis of CpG islands in four fish genomes. Comp Funct Genom 2008. doi: 10.1155/2008/565631.

Inácio A, Pinho J, Pereira PM, Comai L, Coelho MM (2012). Global analysis of the small RNA transcriptome in different ploidies and genomic combinations of a vertebrate complex – the Squalius alburnoides. PLoS One 7: e41158.

Katagiri T, Asakawa S, Hirono I, Aoki T, Shimizu N (2000). Genomic bacterial artificial chromosome library of the Japanese flounder Paralichthys olivaceus. Mar Biotechnol 2: 571–576.

Kawaguchi M, Hiroi J, Miya M, Nishida M, Iuchi I, Yasumasu S (2010). Intron-loss evolution of hatching enzyme genes in Teleostei. BMC Evol Biol 10: 260.

Koroma AP, Jones R, Michalak P (2011). Snapshot of DNA methylation changes associated with hybridization in Xenopus. Physiol Genomics 43: 1276–1280.

Liu S (2010). Distant hybridization leads to different ploidy fishes. Sci China Life Sci 53: 416–425.

Liu S, Liu Y, Zhou G, Zhang X, Luo C, Feng H et al. (2001). The formation of tetraploid stocks of red crucian carp×common carp hybrids as an effect of interspecific hybridization. Aquaculture 192: 171–186.

Ma X-F, Gustafson JP (2006). Timing and rate of genome variation in triticale following allopolyploidization. Genome 49: 950–958.

Mallet J (2007). Hybrid speciation. Nature 446: 279–283.

McClintock B (1984). The significance of responses of the genome to challenge. Science 226: 792–801.

Moss SP, Joyce DA, Humphries S, Tindall KJ, Lunt DH (2011). Comparative analysis of teleost genome sequences reveals an ancient intron size expansion in the zebrafish lineage. Genome Biol Evol 3: 1187–1196.

Otto SP (2007). The evolutionary consequences of polyploidy. Cell 131: 452–462.

Pala I, Coelho MM, Schartl M (2008). Dosage compensation by gene-copy silencing in a triploid hybrid fish. Curr Biol 18: 1344–1348.

Salamov AA, Solovyev VV (2000). Ab initio gene finding in Drosophila Genomic DNA. Genome Res 10: 516–522.

Salmon A, Ainouche ML, Wendel JF (2005). Genetic and epigenetic consequences of recent hybridization and polyploidy in Spartina (Poaceae). Mol Ecol 14: 1163–1175.

Salmon A, Flagel L, Ying B, Udall JA, Wendel JF (2010). Homoeologous nonreciprocal recombination in polyploid cotton. New Phytol 186: 123–134.

Sehrish T, Symonds V, Soltis D, Soltis P, Tate J (2014). Gene silencing via DNA methylation in naturally occurring Tragopogon miscellus (Asteraceae) allopolyploids. BMC Genomics 15: 701.

Song K, Lu P, Tang K, Osborn TC (1995). Rapid genome change in synthetic polyploids of Brassica and its implications for polyploid evolution. Proc Natl Acad Sci USA 92: 7719–7723.

Tamura K, Peterson D, Peterson N, Stecher G, Nei M, Kumar S (2011). MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. Mol Biol Evol 28: 2731–2739.

Thompson JD, Higgins DG, Gibson TJ (1994). CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. Nucleic Acids Res 22: 4673–4680.

Ungerer MC, Strakosh SC, Zhen Y (2006). Genome expansion in three hybrid sunflower species is associated with retrotransposon proliferation. Curr Biol 16: R872–R873.

Venkatesh B, Ning Y, Brenner S (1999). Late changes in spliceosomal introns define clades in vertebrate evolution. Proc Natl Acad Sci USA 96: 10267–10271.

Xiao J, Song C, Liu S, Tao M, Hu J, Wang J et al. (2013). DNA methylation analysis of allotetraploid hybrids of red crucian carp (Carassius auratus red var.) and common carp (Cyprinus carpio L.). PLoS One 8: e56409.

Xu P, Li J, Li Y, Cui R, Wang J, Wang J et al. (2011). Genomic insight into the common carp (Cyprinus carpio) genome by sequencing analysis of BAC-end sequences. BMC Genomics 12: 188.

Zou J, Fu D, Gong H, Qian W, Xia W, Pires JC et al. (2011). De novo genetic variation associated with retrotransposon activation, genomic rearrangements and trait variation in a recombinant inbred line population of Brassica napus derived from interspecific hybridization with Brassica rapa. Plant J 68: 212–224.

Supplementary Information accompanies this paper on Heredity website (http://www.nature.com/hdy)